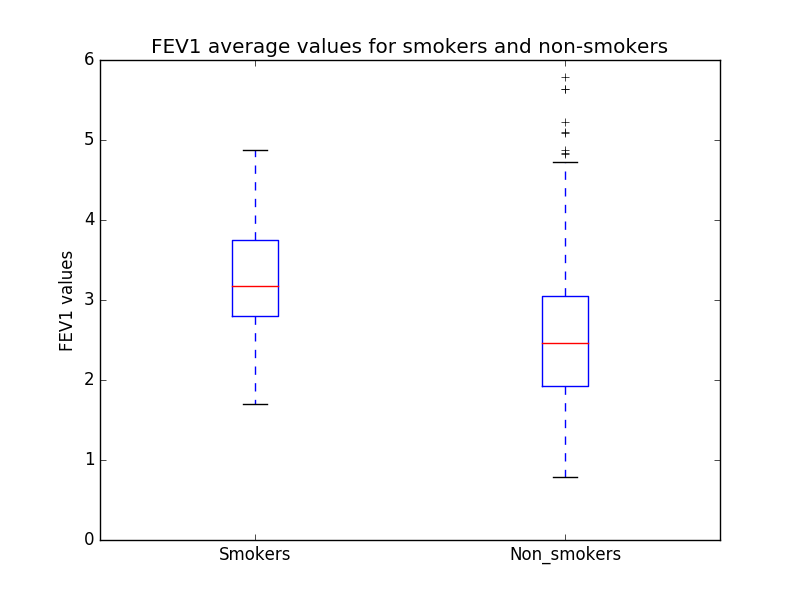
**Assignment 1 – Introduction to Data Science**

**Exercise 1**

The average FEV1 values for smokers and non-smokers are: (3.2768615384615383, 2.5661426146010187).

The values show a lower average FEV1 for non-smokers which should indicates a derease in lung function. This doesn’t match with what was expected: the FEV1 should be lower for the smokers group as it’s widely known that smoking affects the breathing.

**Exercise 2**

****

The plot shows that the FEV1 values are higher for the smokers group that for the non-smokers. The red line represents the median of the distribution. In the non-smokers data, some cross-hatches with high values of FEV1 are plotted even though the median results lower than the smokers median.

**Exercise 3**

The value of the t-statistic is: 6.4644531726

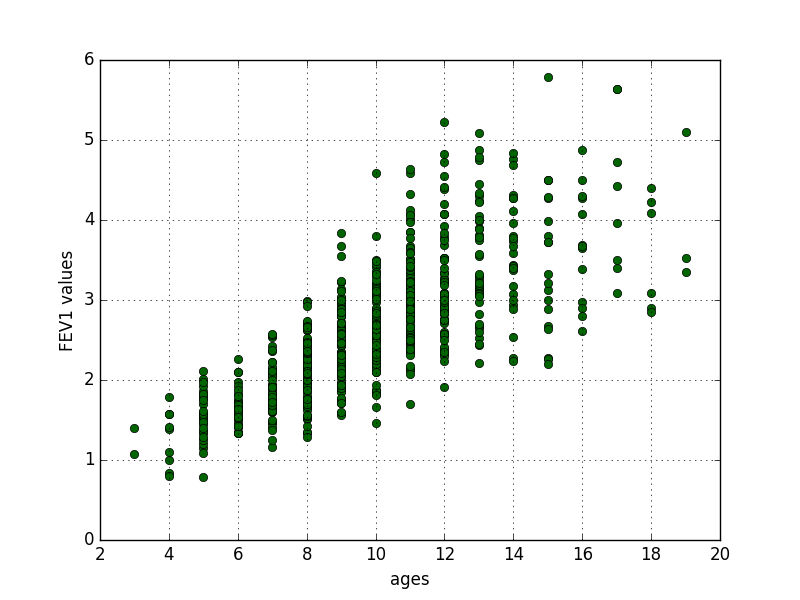
Degrees of freedom of value: is the sum of the samples sizes minus 2. In our case, the sum of the samples is the size of our original dataset, which is equal to 654 so the degrees of freedom is v=654-2=652.

The p value of the samples is: 1.99284591829e-10

So the null hypothesis is rejected because the p value is smaller than the significance level.

#TODO short discussion of the result

**Exercise 4**

****

The Pearson correlation coefficient, and the p-value are: (0.75645898998959982, 2.4539601903501647e-122)

**Exercise 5**

